

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе и
образовательным инновациям

О.Н.Здрок

«30» сентября 2020 г.

Регистрационный № УД- 9244/уч.

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности**

1-31 80 09 Прикладная математика и информатика

Профилизация: Интеллектуальные системы

2020г.

Учебная программа составлена на основе ОСВО 1-31 80 09-2019 и учебного плана G31-128/уч. от 11.04.2019 г.

СОСТАВИТЕЛИ:

Н.К. Рубашко – старший преподаватель кафедры информационных систем управления факультета прикладной математики и информатики Белорусского государственного университета

РЕЦЕНЗЕНТЫ:

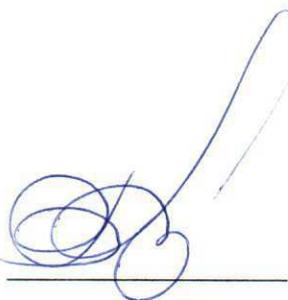
А.А. Дудкин – заведующий лабораторией идентификации систем ОИПИ НАН Беларуси, доктор технических наук, профессор.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой информационных систем управления
(протокол № 9 от 20 марта 2020 года);

Научно-методическим Советом БГУ
(протокол № 4 от 25 марта 2020 года).

Заведующий кафедрой
информационных систем управления



В.В. Краснопрошин

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Учебная дисциплина «Компьютерная лингвистика» (КЛ) знакомит студентов магистратуры с теоретическими положениями лингвистического и математического моделирования естественного языка, с вариантами технологических решений, обеспечивающих компьютерную форму существования естественного языка. Дается лингвистическая интерпретация основных лингвистических объектов и единиц анализа, рассматриваются общие вопросы моделирования лингвистических сущностей. Дисциплина базируется на современных достижениях в области обработки естественного языка.

Цель учебной дисциплины – формирование у студентов магистратуры навыков системного представления о лингвистических основах информатики и аспектах моделирования языка и мышления в компьютерной среде.

Задачи учебной дисциплины:

1. создание теоретического базиса для использования современных технологий обработки данных при решении задач компьютерной лингвистики;
2. развитие у магистрантов навыков по проведению исследовательской работы, овладения ими методологией анализа результатов решения как теоретических, так и практических задач.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием (магистра).

Учебная дисциплина относится к модулю «Основы искусственного интеллекта» компонента учреждения высшего образования.

Программа составлена с учетом **межпредметных связей** с учебными дисциплинами. Основой для изучения являются учебные дисциплины первой ступени высшего образования по дискретной математике и математической логике, теории алгоритмов, моделям данных и системам управления базами данных и дисциплина второй ступени высшего образования «Модели и методы искусственного интеллекта».

Требования к компетенциям

Освоение учебной дисциплины «Компьютерная лингвистика» должно обеспечить формирование следующих специализированных и углубленных профессиональных компетенций:

специализированные компетенции:

СК–7. Уметь ставить и решать прикладные задачи компьютерной лингвистики, определять методы и средства их эффективного решения;

СК–13. Уметь использовать научные и технические достижения для разработки эффективных алгоритмов решения прикладных задач;

углубленные профессиональные компетенции:

УПК–5. Владеть перспективными технологиями программирования.

В результате изучения дисциплины студент магистратуры должен:

знать:

- основные понятия, составляющие базу современной лингвистической науки, и ее термины;
- методы и решения в системах организации знаний; эмпирические законы описания естественных языков;
- основные задачи и области применения компьютерной лингвистики;
- связь компьютерной лингвистики с другими дисциплинами прикладной математики и лингвистики;

уметь:

- пользоваться основными понятиями прикладной математики и лингвистики, применяемыми в области компьютерной лингвистики;
- реализовывать различные алгоритмы обработки естественного языка с использованием компьютера;

владеть:

- знаниями о системе языковедческих дисциплин; предмете, методах и моделях прикладной лингвистики; предмете компьютерной лингвистики;
- основными методами и приемами исследовательской и практической работы в области лингвистики.

Структура учебной дисциплины

Дисциплина изучается во 2 семестре. Всего на изучение учебной дисциплины «Компьютерная лингвистика» отведено:

– для очной формы получения высшего образования – 126 часов, в том числе 40 аудиторных часов, из них: лекции – 20 часов, лабораторные занятия – 20 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма текущей аттестации по учебной дисциплине – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Краткое введение в проблематику КЛ

Тема 1.1. Введение в предмет КЛ

Объект исследования современной лингвистики текста. Информатика и компьютерная лингвистика. Область возникновения лингвистических проблем информатики. Основы теории речевой коммуникации. Совершенствование массовой и индивидуальной коммуникации. Особенности системы естественного языка (ЕЯ): уровни и связи.

Тема 1.2. Компьютерная лингвистика: методы, ресурсы, приложения

Характеризация классов задач компьютерной лингвистики. Приложения компьютерной лингвистики. WRE как обобщение регулярных выражений. Применение в компьютерной лингвистике

Раздел 2. Моделирование в компьютерной лингвистике

Тема 2.1. Моделирование языковых сущностей и человеческого мышления

Связь языка с мышлением. Элементы системы искусственного интеллекта. Модель механизма мышления. Ассоциативное построение понятий. Основные принципы работы системы и организации ее поведения.

Тема 2.2. Знание как объект моделирования

Представление знаний. Семантические сети и фреймы. Лингвистический аспект представления знаний. Понимание текста. Денотативный анализ текста.

Тема 2.3. Моделирование обучения языку

Обучение языку. Сравнение моделей.

Тема 2.4. Моделирование на уровне статистики

Квантитативная лингвистика и связанные с ней дисциплины. Теория и результаты. Язык и речь. Аспекты речевой деятельности.

Раздел 3. Лингвистические ресурсы

Тема 3.1. Историческое развитие слова и его значений

Лексика современных текстов. Развитие терминологической лексики. Основной словарный состав. Выбор слов исходного массива. Выбор словарей.

Тема 3.2. Машиночитаемые словари

Традиционная и машинная лексикография. Отличия машинного словаря от обычного. Лемматизация. Составление машинных словарей. Контекстологический словарь как организатор базы знаний.

Тема 3.3. Корпус текстов и корпусная лингвистика

Основные понятия корпусной лингвистики. Основные характеристики корпусов. Проектирование и технологический процесс создания.

Стандартизация в корпусной лингвистике. Корпус как поисковая система. Обзор существующих корпусов различных типов. Корпусные исследования.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования с применением дистанционных образовательных технологий

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов			Форма контроля знаний
		Лекции	Лабораторные занятия	Количество часов УРС	
1	2	3	4	5	6
1	КРАТКОЕ ВВЕДЕНИЕ В ПРОБЛЕМАТИКУ КЛ	4			Устный опрос.
1.1	Введение в предмет КЛ	2			
1.2	Компьютерная лингвистика: методы, ресурсы, приложения	2			
2	МОДЕЛИРОВАНИЕ В КОМПЬЮТЕРНОЙ ЛИНГВИСТИКЕ	6	8		Устный опрос. Отчет по лабораторным работам. Выполнение тестов. Контрольная работа 1
2.1	Моделирование языковых сущностей и человеческого мышления	2	2		
2.2	Знание как объект моделирования	1	2		
2.3	Моделирование обучения языку	1	2		
2.4	Моделирование на уровне статистики	2	2		
3	ЛИНГВИСТИЧЕСКИЕ РЕСУРСЫ	10	12		Устный опрос. Отчет по лабораторным работам. Выполнение тестов. Контрольная работа 2
3.1	Историческое развитие слова и его значений	2			
3.2	Машиночитаемые словари	4	6		
3.3	Корпус текстов и корпусная лингвистика	4	6		
	ВСЕГО:	20	20		

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Прикладная и компьютерная лингвистика 2016. – Москва: Ленард, 2016. – 320 с.
2. Баркович А.А. Информационная лингвистика: Метаописания современной коммуникации. - Москва : ФЛИНТА, 2017. - 358 с.
3. Материалы международной конференции по компьютерной лингвистике и интеллектуальным технологиям Dialog 2008-2021 [Электронный ресурс]. – Режим доступа: <http://www.dialog-21.ru/>
4. Филиппович Ю.Н. Лингвистическое обеспечение информационных систем. Часть 1. Компьютерная лингвистика. Начало (посл. четв. XX века). Учебное пособие. М.: МГУП, 2013. – 452 с.
5. Автоматическая обработка текста на естественном языке и компьютерная лингвистика: учебное пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Песков О.В., Ягунова Е.В. – М.: МИЭМ, 2011.
6. Марчук Ю.Н. Компьютерная лингвистика. – М.: Изд-во АСТ. 2007 г. – 320 с.
7. Баранов А.Н. Введение в прикладную лингвистику. Изд.4, испр. и доп. –М.: УРСС, 2013. 3– 68 с.
8. Jurafsky, D. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition / D. Jurafsky, J. H. Martin. – New Jersey: Prentice Hall PTR, 2000. – 934 p.

Перечень дополнительной литературы

1. Овчинникова И.Г., Угланова И.А. Компьютерное моделирование вербальной коммуникации 2009. 136 с.
2. Всеволодова А.В. Компьютерная обработка лингвистических данных. Изд.2 2007. 96 с.
3. Ландэ Д.В., Снарский А.А., Безсуднов И.В. Интернетика: Навигация в сложных сетях: модели и алгоритмы, 2009.
4. Fastus: A cascaded finite-state transducer for extracting information from natural-language text / D. Israel [et al.] // Finite State Devices for Natural Language Processing / ed. by Roche, Schabes. – Cambridge, MA, USA: MIT Press, 1996. – P. 383–406.
5. CLAIM – научно-образовательный кластер. Серия книг "Компьютерная лингвистика". – Режим доступа: <http://it-claim.ru/Library/Books/CL/CLbook.htm>

6. Литература по компьютерной лингвистике". – Режим доступа: <http://www.compling-info.narod.ru>
7. Международная конференция Диалог. – Режим доступа: Dialog-21.ru
8. Лаборатория компьютерной лингвистики Института проблем передачи информации РАН. – Режим доступа: Philol.msu.ru/~lex/library.htm

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

Для диагностики компетенции в рамках учебной дисциплины рекомендуется использовать следующие формы:

1. Устная форма: устный опрос, коллоквиум, выступление с докладом на семинаре.
2. Письменная форма: контрольные работы.
3. Устно-письменная форма: отчеты по домашним практическим заданиям с их устной защитой, оценивание на основе проектного метода, выполнение тестов.

Формой текущей аттестации по дисциплине «Компьютерная лингвистика» учебным планом предусмотрен зачет.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний студентов по дисциплине.

Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний в рейтинговую оценку (формирование оценки за текущую успеваемость):

- отчёты по лабораторным работам – 35 %;
- контрольные работы – 30 %;
- тесты – 25%;
- устный опрос – 10%.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и зачетной оценки с учетом их весовых коэффициентов. Вес оценки по текущей успеваемости составляет 30 %, зачетной оценки – 70 %.

Примерная тематика лабораторных занятий

Задание № 1. Создание частотного индекса на заданном корпусе текстов.

Задание № 2. Разработка лексической онтологии заданного корпуса текстов.

Задание № 3. Визуализация онтологии.

Задание № 4. Разработка машинных словарей.

Задание № 5. Разработка терминологических словарей.

Задание № 6. Разработка особых типов словарей.

Рекомендуемая тематика контрольных работ

Контрольная работа №1. Моделирование на уровне статистики.

Контрольная работа №2. Технологии создания машинных словарей.

Текущий контроль знаний проводится в соответствии с учебно-методической картой дисциплины.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса большинства практических занятий используется *практико-ориентированный подход*, который предполагает освоение содержания учебного материала через решение практических задач, а также приобретение навыков эффективного выполнения разных видов профессиональной деятельности.

Кроме этого, при организации образовательного процесса используется комбинация таких методов *креативного обучения*, как *методы группового обучения, проектного обучения и учебной дискуссии*. Комбинация методов предполагает ориентацию на генерирование идей, приобретение навыков для решения исследовательских, творческих и коммуникационных задач, появление нового уровня понимания изучаемой темы, применение знаний (теорий, концепций) при решении проблем, определение способов их решения.

Методические рекомендации по организации самостоятельной работы обучающихся, подготовка к экзамену

Для организации самостоятельной работы студентов магистратуры по учебной дисциплине следует использовать информационно-коммуникационные технологии:

- образовательный портал БГУ <https://edufpmi.bsu.by>;
- образовательный портал InsightRunner <https://acm.bsu.by>;
- систему AnyTask <https://anytask.org/school/bsu>;

разместить в сетевом доступе комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, презентации лекций, методические указания к практическим занятиям, электронные версии домашних заданий, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в том числе вопросы для подготовки к экзамену, задания,

вопросы для самоконтроля, список рекомендуемой литературы, информационных ресурсов и др.).

Примерный перечень вопросов к зачету

1. Предмет компьютерной лингвистики.
2. История становления и основные тенденции развития.
3. Основные направления компьютерной лингвистики.
4. Компьютерное обеспечение представления знаний.
5. Автоматизированный анализ: распознавание и синтез устной и письменной речи.
6. Морфологический анализ, проблемы семантического анализа, синтаксический анализ.
7. Лингвистические базы данных: модели и типы данных.
8. Создание общих искусственных языков для представления информации.
9. Компьютерная лексикография как одно из направлений прикладной лингвистики.
10. Словарные процессоры.
11. Основные понятия структуры словаря: словник, словарная статья, грамматические, стилистические пометы; иллюстративный материал
12. Типология электронных словарей.
13. Тезаурусы и терминологические словари.
14. Компьютерные технологии составления и эксплуатации словарей.
15. Специальные программы – базы данных, программы обработки текста.
16. Корпусная лингвистика, ориентированная на создание национальных корпусов естественных языков.
17. Типы лингвистической разметки.
18. Концепция обучения на основе компьютерных технологий.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Интеллектуальные системы мониторинга	Информационных систем управления	Нет	Оставить содержание учебной дисциплины без изменения, (протокол № 9 от 20 марта 2020 г.)

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ

на ____ / ____ учебный год

№№ Пп	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры информационных систем управления (протокол № ____ от _____ 20__ г.)

Заведующий кафедрой

Д.т.н., профессор

(ученая степень, звание)

(подпись)

В.В.Краснопрошин

(И.О. Фамилия)

УТВЕРЖДАЮ

Декан факультета

Д.т.н., доцент

(ученая степень, звание)

(подпись)

А.М. Недзведзь

(И.О.Фамилия)